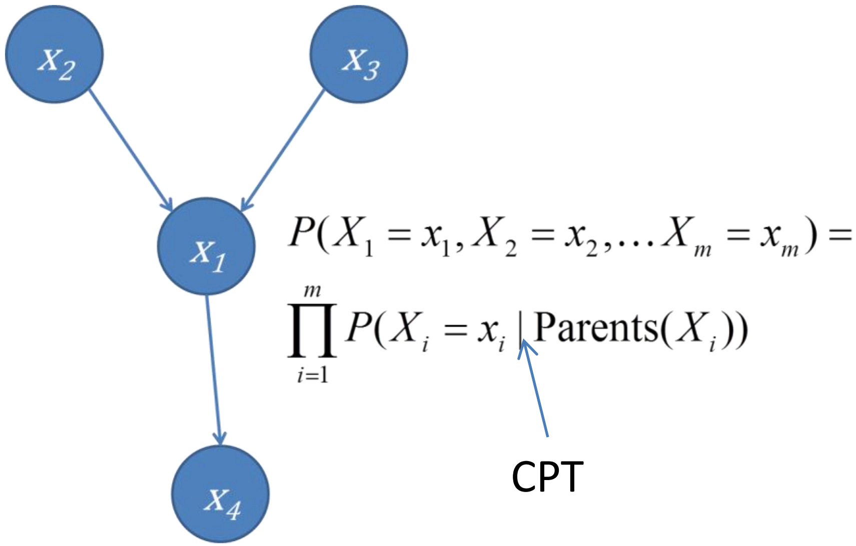
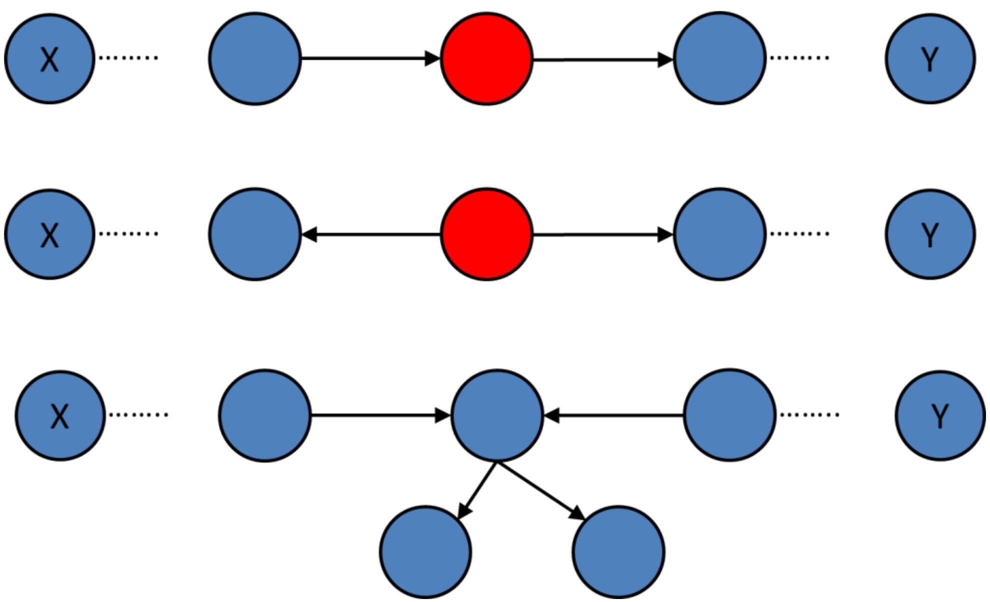
Reasoning with uncertainty II

Summary



$$P(X_1 = x_1, X_2 = x_2, \ldots X_m = x_m) = \prod_{i=1}^{m} P(X_i = x_i \mid \text{Parents}(X_i))$$

Exact inference linear in the number of nodes ($d^{k+1}n$) for polytrees

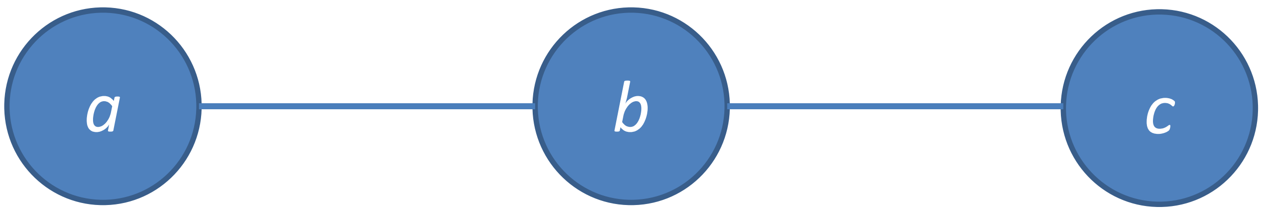


Graphical tests for independence structure

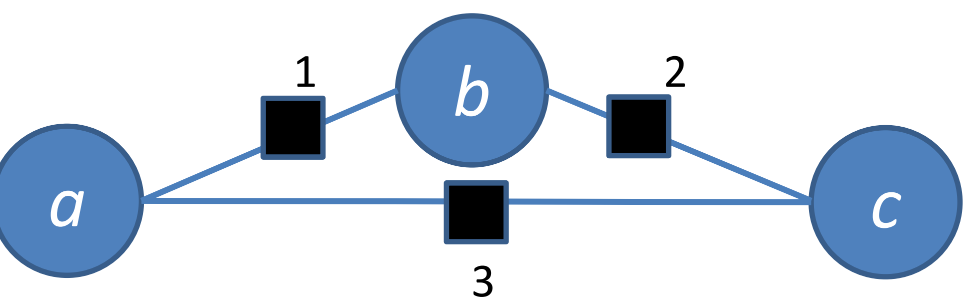
Summary: Undirected

$$P(a,b) = \varphi(a,b)$$

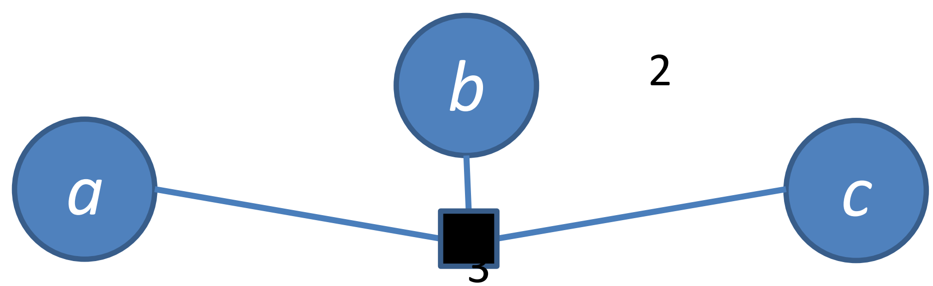
$$P(a,b) = \varphi_1(a)\varphi_2(b)$$



$$P(a,b,c) = \varphi_1(a,b)\,\varphi_2(b,c) \quad a \perp c \mid b$$

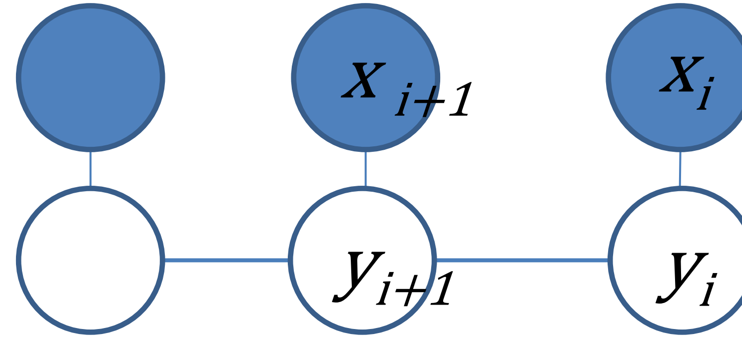


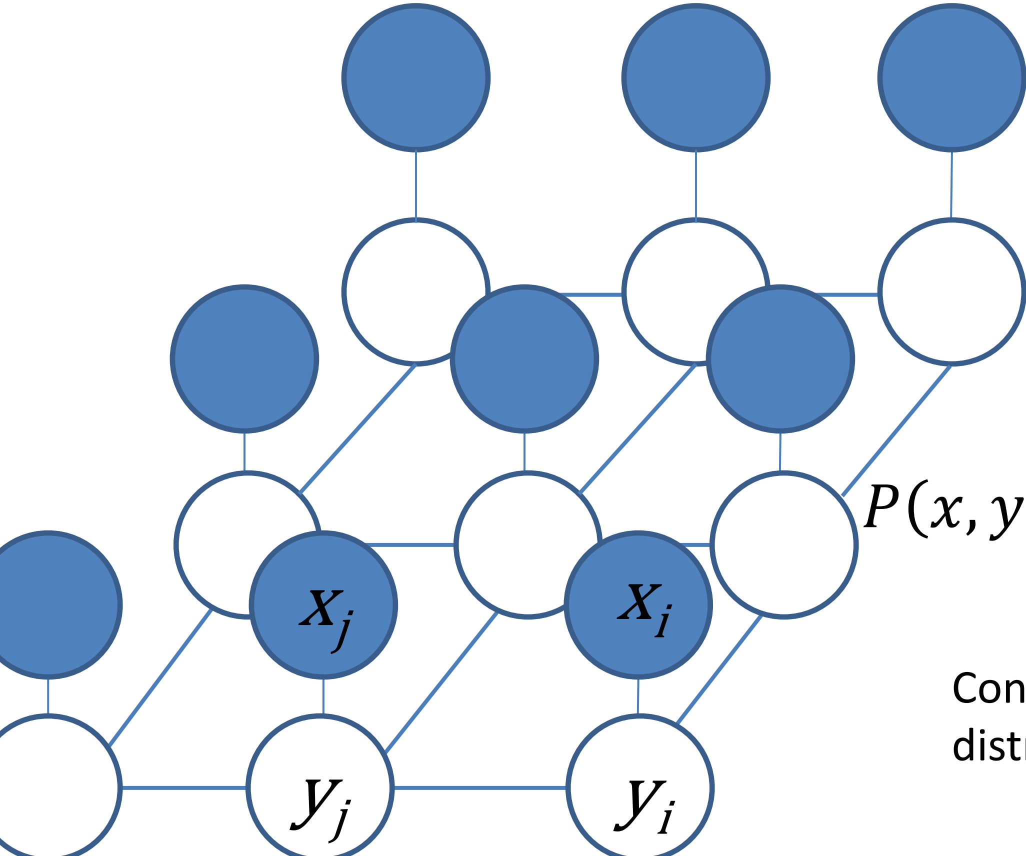
$$P(a,b,c) = \varphi_1(a,b)\,\varphi_2(b,c)\,\varphi_3(a,c)$$



$$P(a,b,c) = \varphi\,(a,b,c)$$

Example





$$P(x,y)\ \alpha \prod_i \psi_i(y_i, y_{i+1}) \prod_i \varphi_i(x_i, y_i)$$

Consistency on spatial distribution of labels

Agreement of label vs. input data

$$P(x,y)\ \alpha \prod_{i,j} \varphi_{ij}(y_i, y_j) \prod_i \varphi_i(x_i, y_i)$$

- y = decision variable (class label (e.g., road, car, etc…))
- x = observation variable (e.g., image patch)

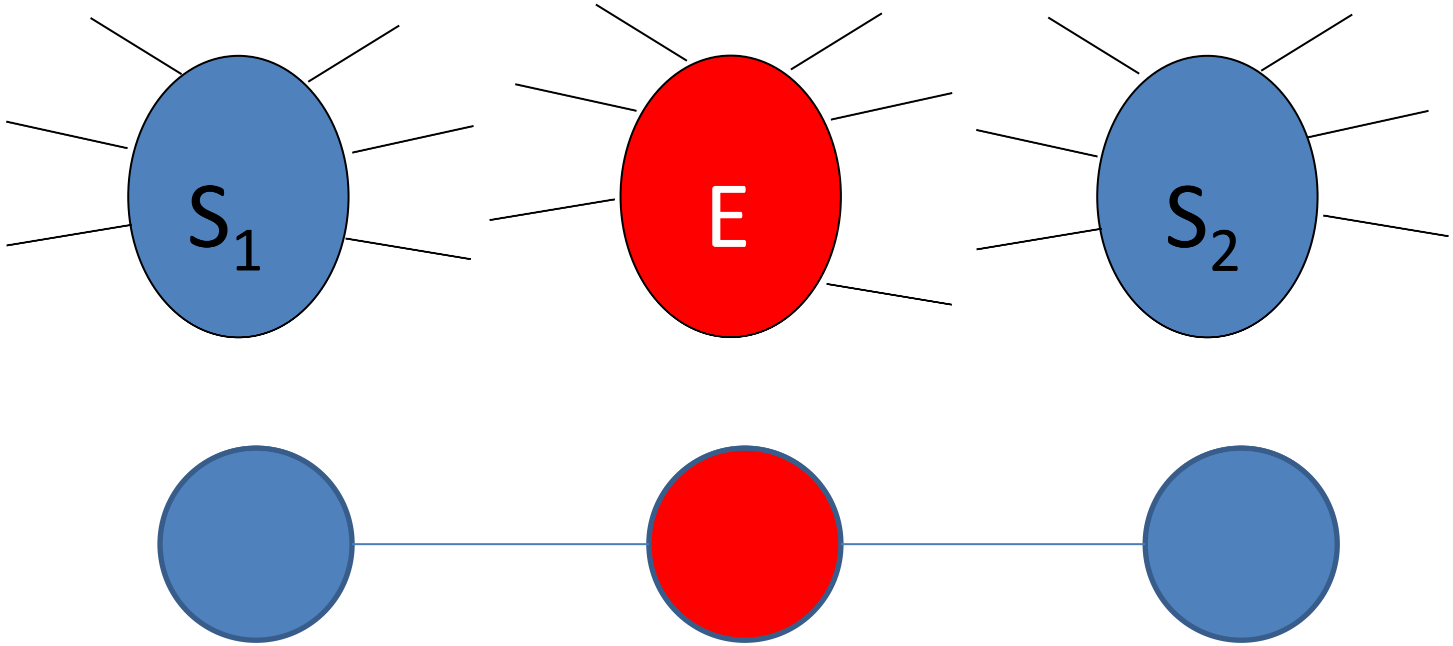
Undirected: What factors

- Factors do not need to be CPTs or combinations of probabilities
- Only condition: non-negative
- Intuitively: Potential measuring compatibility between nodes
- But need normalization

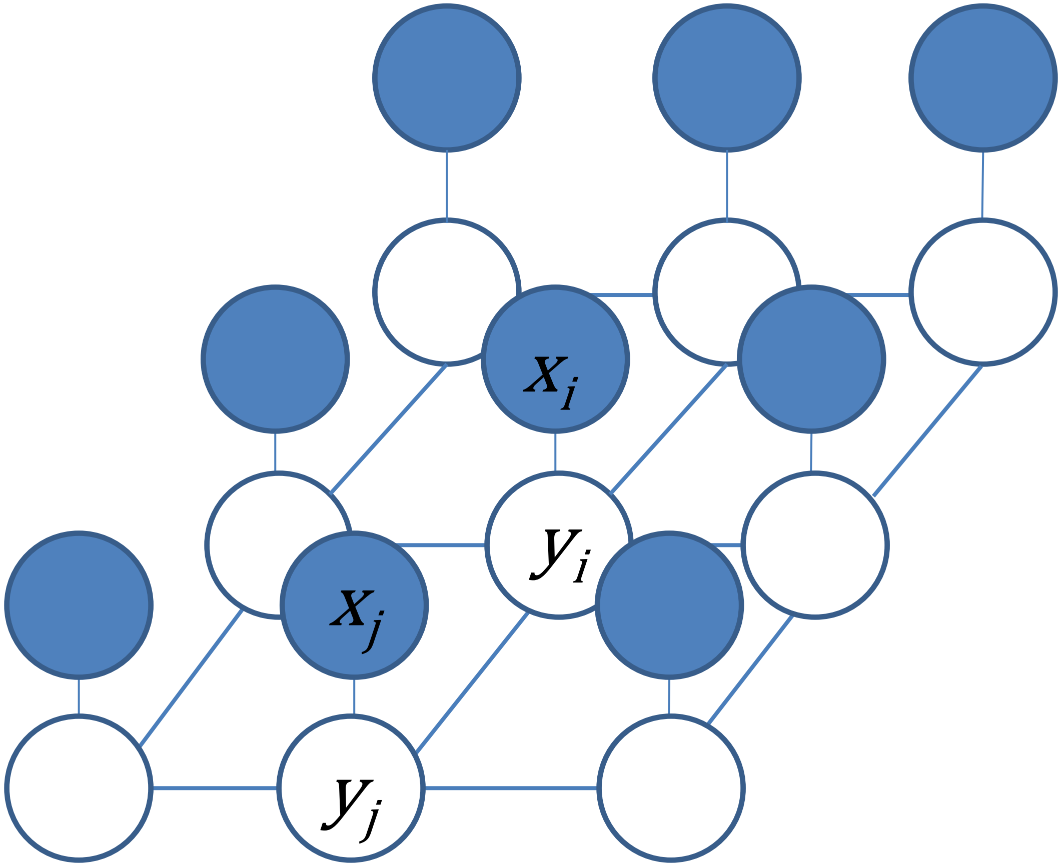
$$P(x_1, x_2, .., x_n) = \frac{1}{Z} \prod_A \varphi_A(x_{A1}, .., x_{Ak})$$

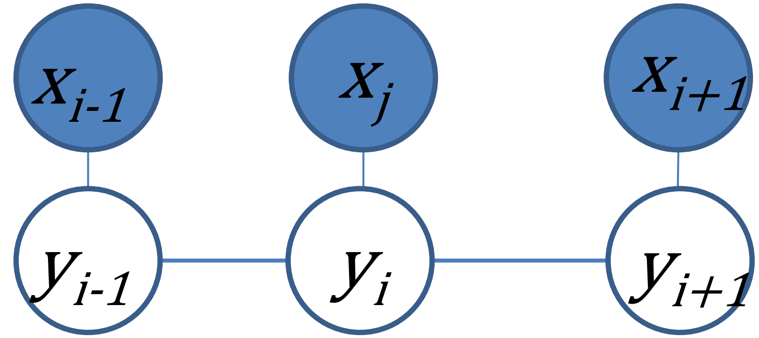
Independence

- Simpler condition:



- If all paths are blocked by E then $S_1 \perp S_2 \mid E$
- Purely graphical property independent of actual φ's

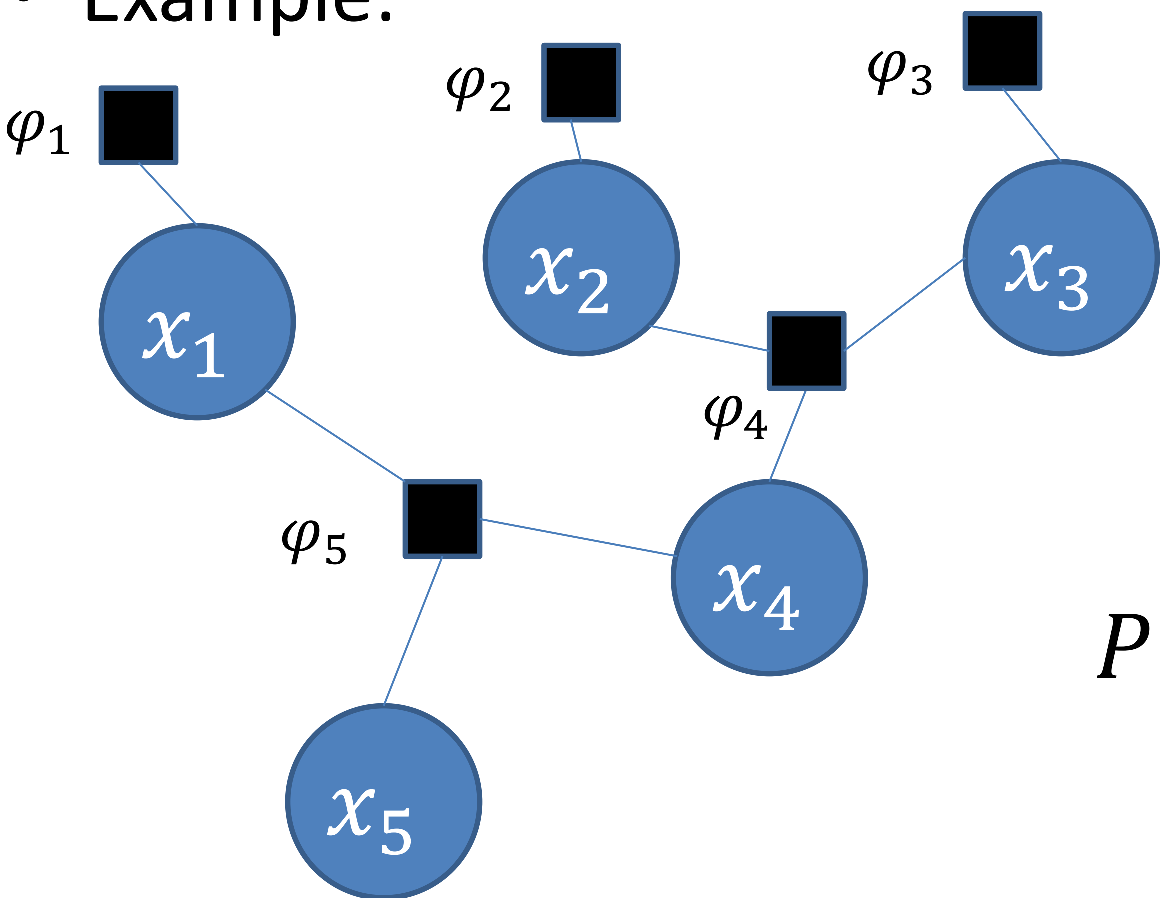




- y = decision variable (class label (e.g., road, car, etc…))
- x = observation variable (e.g., image patch)
- Markov property of local dependence

Inference

- General query probability distribution of set of variables E_1given set values of other set E_2
- In general: $P(E_1|E_2)\alpha\sum_{E_3}P(x_1,..,x_n)$
- Example:



$$P(x_4|x_2,x_5)$$

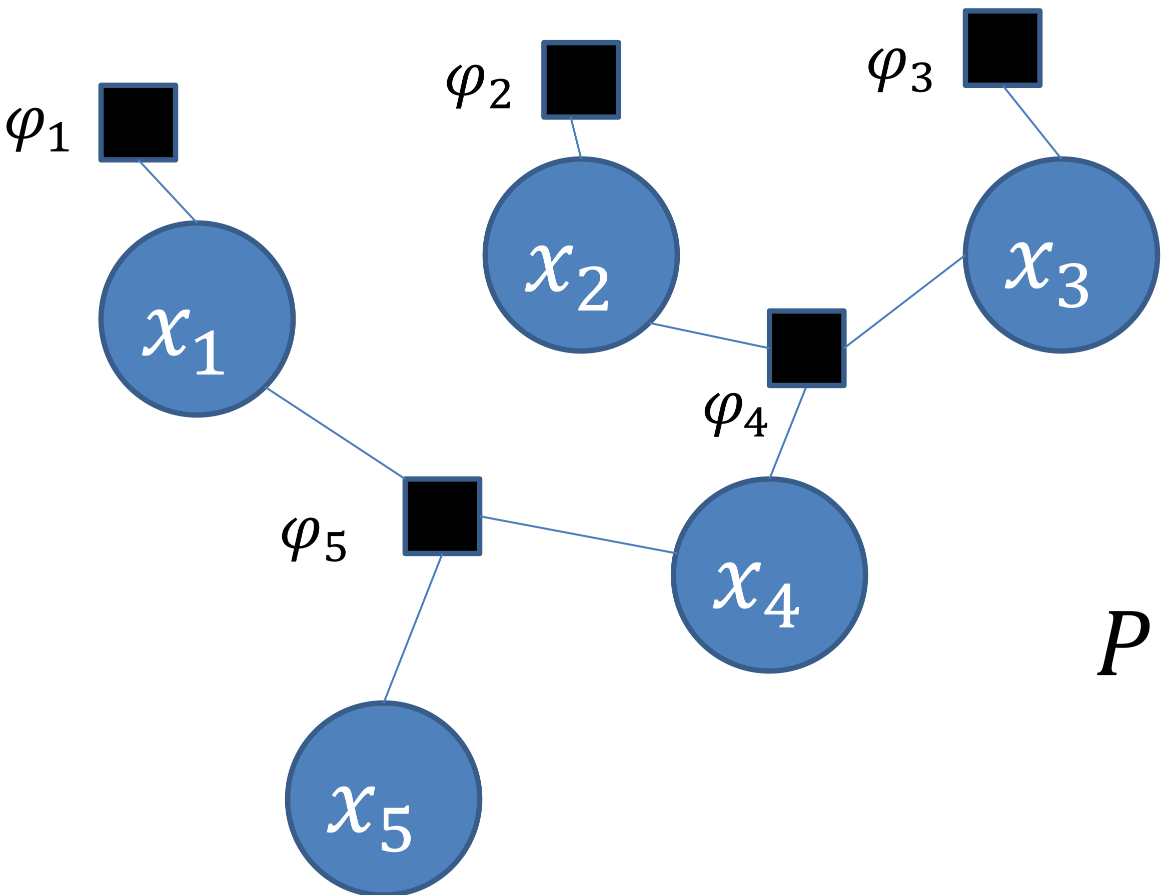
Inference

$P(x_1,..,x_n)\ \alpha\ \varphi_1(x_1)\varphi_2(x_2)\varphi_3(x_3)\varphi_4(x_2,x_3,x_4)\varphi_5(x_1,x_4,x_5)$

Conditioned on x_2, x_5 so fix them

$P(x_1,..,x_n)\ \alpha\ \varphi_1(x_1)\varphi_2(x_2)\varphi_3(x_3)\varphi_4(x_2,x_3,x_4)\varphi_5(x_1,x_4,x_5)$

$P(x_1,..,x_n)\ \alpha\ \varphi_1(x_1)\varphi_3(x_3)\varphi'_4(x_3,x_4)\varphi'_5(x_1,x_4\ )$

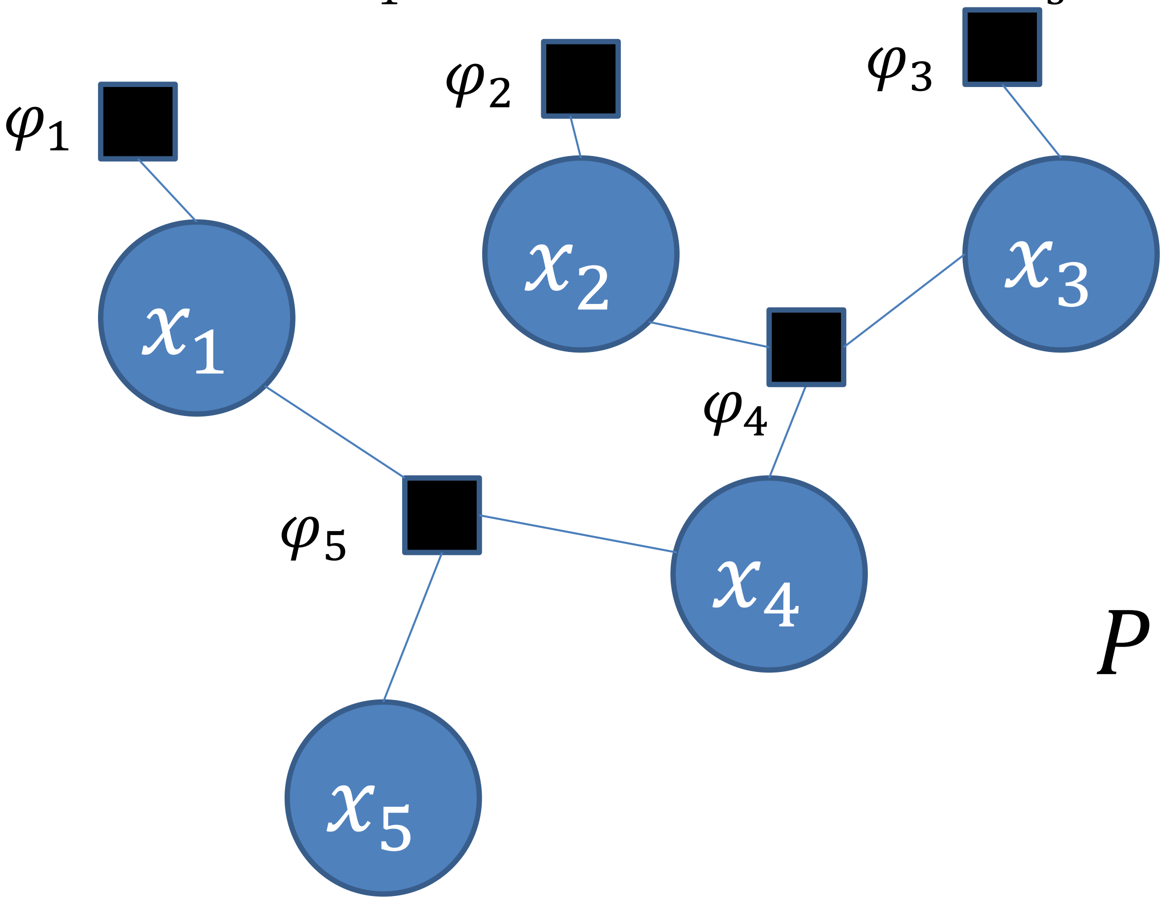


$P(x_4|x_2,x_5)$

Inference

Eliminate the remaining set (E_3) by marginalization

$$P(x_4|x_2,x_5)=\alpha \sum_{x_1} \varphi_1(x_1)\varphi'_5(x_1,x_4) \sum_{x_3} \varphi_3(x_3)\varphi'_4(x_3,x_4)$$



$$P(x_4|x_2,x_5)$$

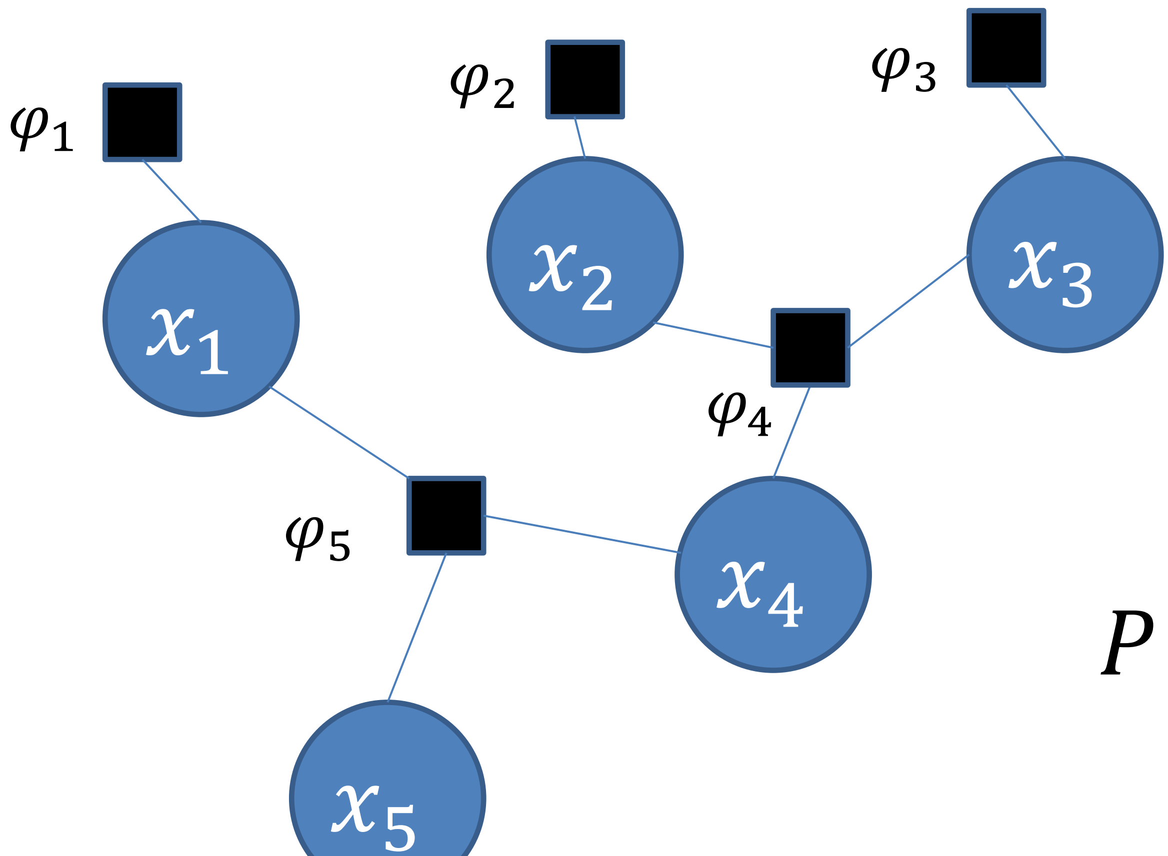
Inference

$$P(x_4|x_2,x_5)=\alpha \sum_{x_1} \varphi_1(x_1)\varphi'_5(x_1,x_4) \sum_{x_3} \varphi_3(x_3)\varphi'_4(x_3,x_4)$$

Finish by normalizing to get a proper probability

Possible because "local" probability

Don't need to normalize before that



$$P(x_4|x_2,x_5)$$

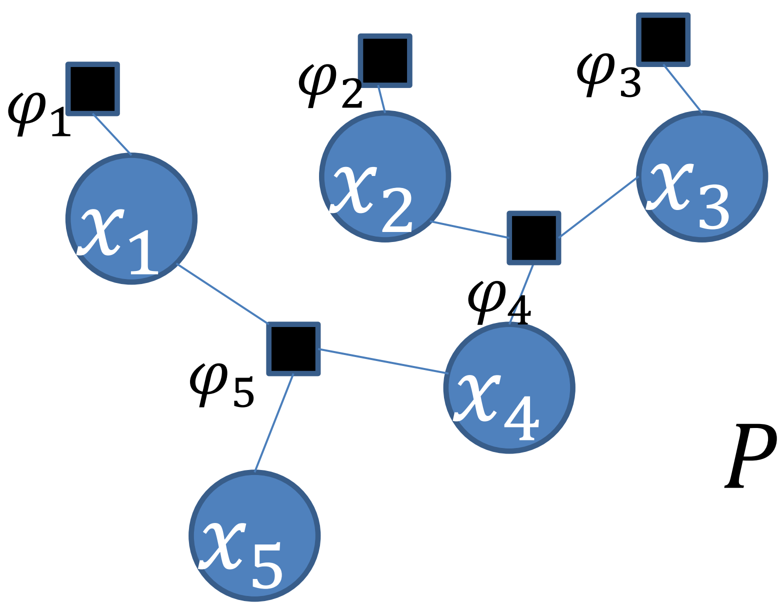
Inference

$$P(x_4|x_2,x_5)=\alpha \sum_{x_1} \varphi_1(x_1)\varphi'_5(x_1,x_4) \sum_{x_3} \varphi_3(x_3)\varphi'_4(x_3,x_4)$$

We were able to group the variables

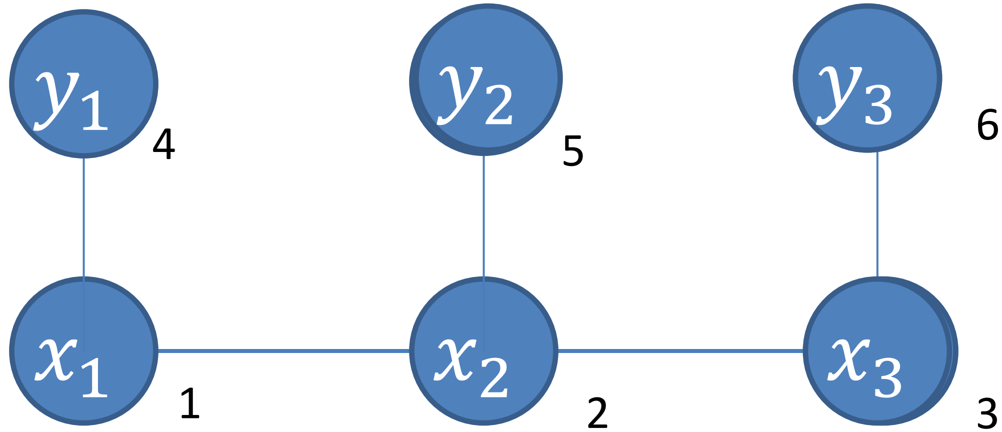
The smaller the group the better

How small can the groups be?

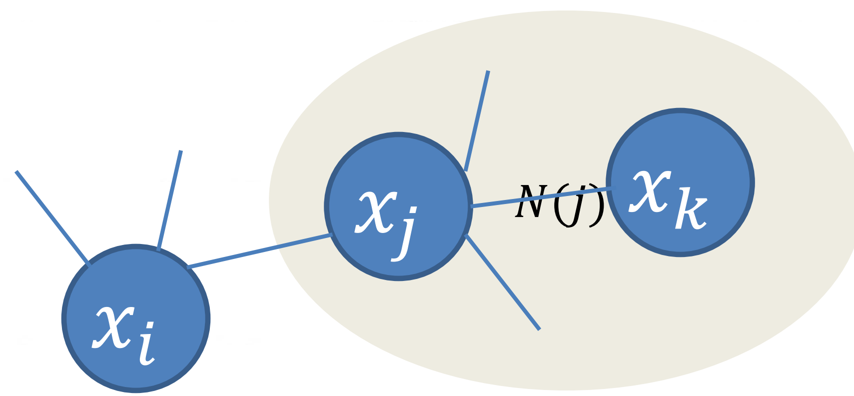


$$P(x_4|x_2,x_5)$$

Example: Tree



$$P(x_1|y) = \sum_{x_2}\sum_{x_3} \phi_{12}(x_1, x_2)\phi_{23}(x_2, x_3)\psi_1(y_1, x_1)\psi_2(y_2, x_2)\psi_3(y_3, x_3)$$



$$m_{ji} = \sum_{x_j} f_{ij}(x_i, x_j) \prod_{N(j)\setminus i} m_{kj}(x_j)$$

- Net result: $|x|^2$ operations instead of $|x|^N$
- General procedure with partial sums:

$$m_{ji} = \sum_{x_j} f_{ij}(x_i, x_j) \prod_{N(j)\setminus i} m_{kj}(x_j)$$

- Take arbitrary node as root
  - Propagate partial sums from leaves
  - Propagate partial sums from root
- Treat factors as nodes, similar approach to passing partial sums

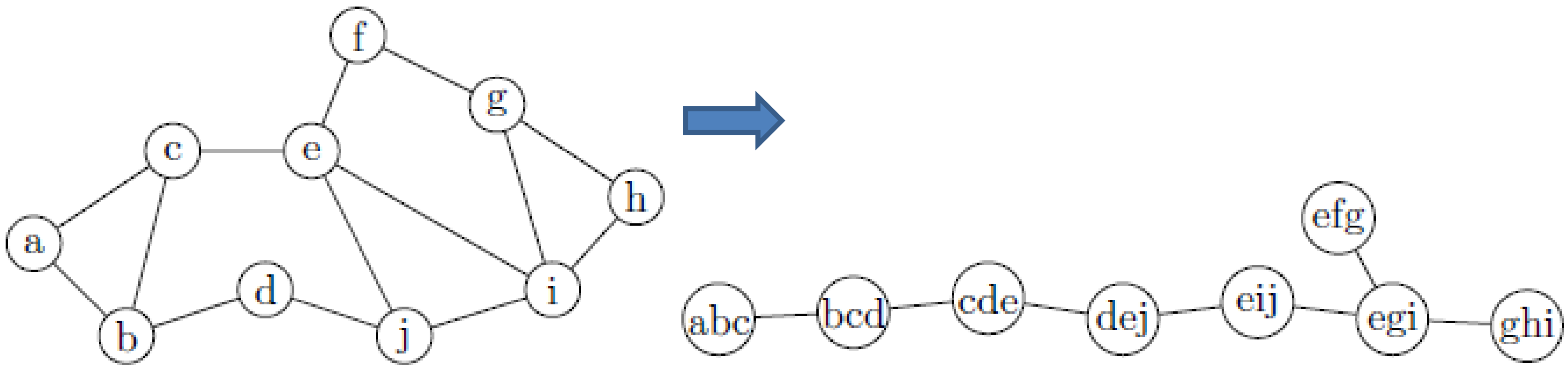
MAP

- Interested in finding $\max P(X)$
- Same idea for distribution applies except with max instead of sum

$$\max_X P(X) = \max_{x_1\, x_2 x_3 x_4} \varphi(x_1, x_2)\varphi(x_2, x_3)\,\varphi(x_3, x_4)$$
$$= \max_{x_1, x_2} \varphi(x_1, x_2) \max_{x_3} \quad \varphi(x_2, x_3) \max_{x_4} \varphi\,(x_3, x_4)$$

$$m_{ji}(x_i) = \max_{x_j} f_{ij}(x_i, x_j) \prod\nolimits_{N(j)\setminus\, i} m_{kj}(x_j)$$

Exact inference?



- Back to the initial question:
  - Exact inference easy on trees (quadratic)
  - Can convert graph to tree with equivalent representation
  - But complexity is size of largest node in the equivalent tree (treewidth+1)
  - Finding the tree with minimum treewidth is NP-hard
  - Approximations, sampling, loopy BP

Connections

- CSP → all values are 0/1 (= constraints between variables; $P(A|B)$ → constraints satisfied when B variables are clamped)
- ILP = MAP assignment

Yanover, Meltzer, Weiss. Linear programming and BP. JML 7.

- Examples:
  – Reasoning about visual and text knowledge

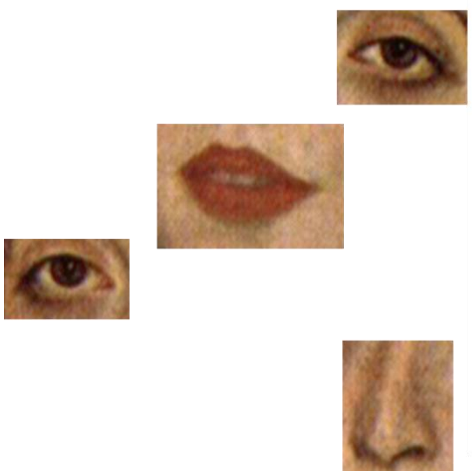
building grass sky

crab rock

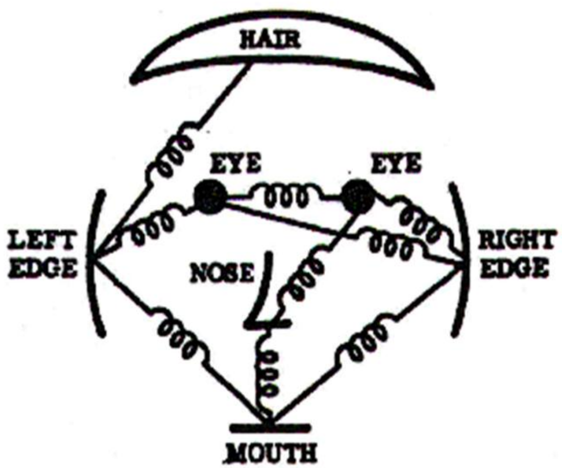
boat water sky house trees

polarbear snow

  – Reasoning about spatial structure



Is this a face?



[Fischler & Elschlager 73]

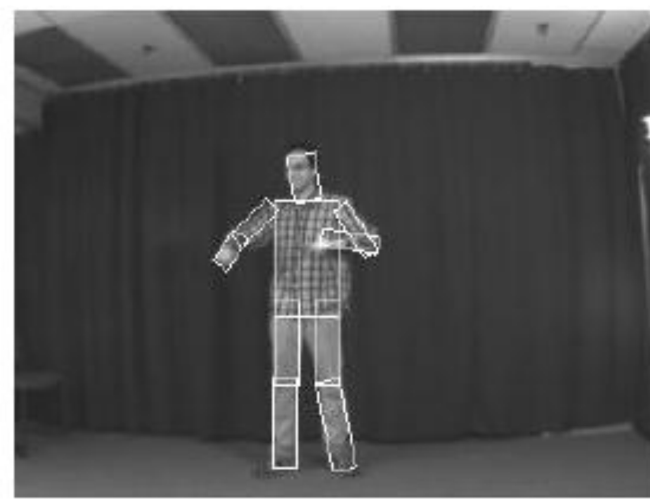
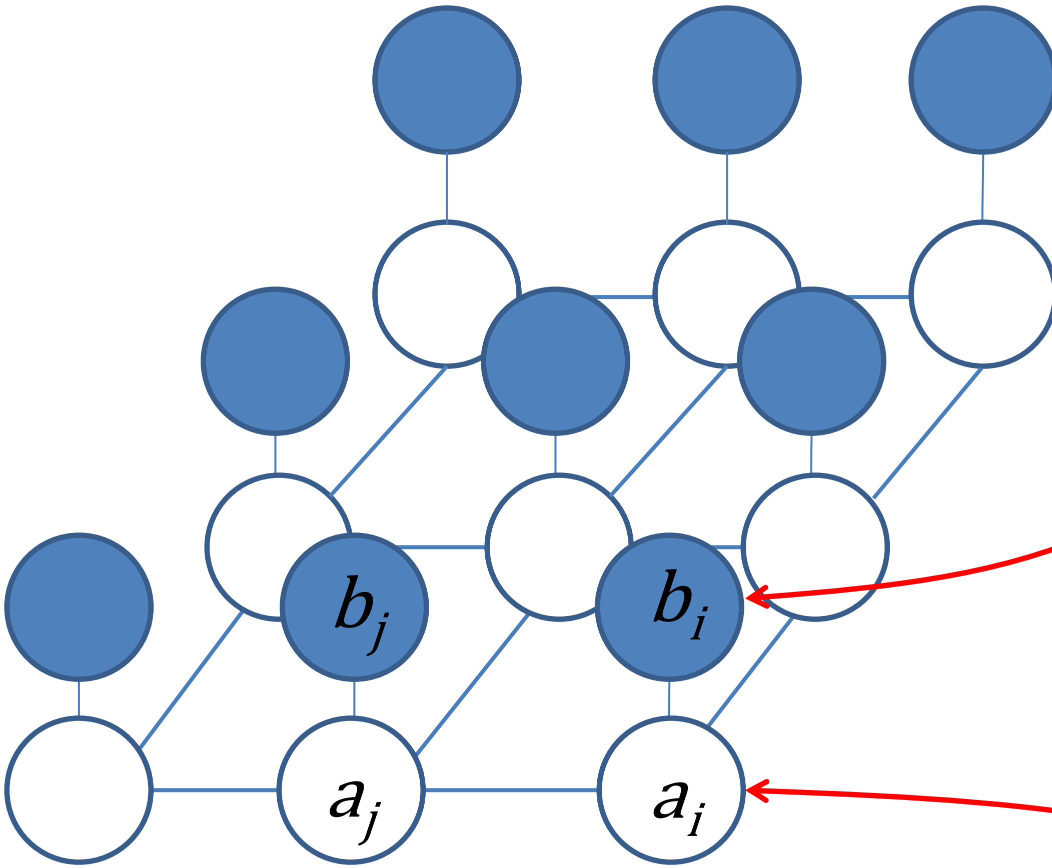


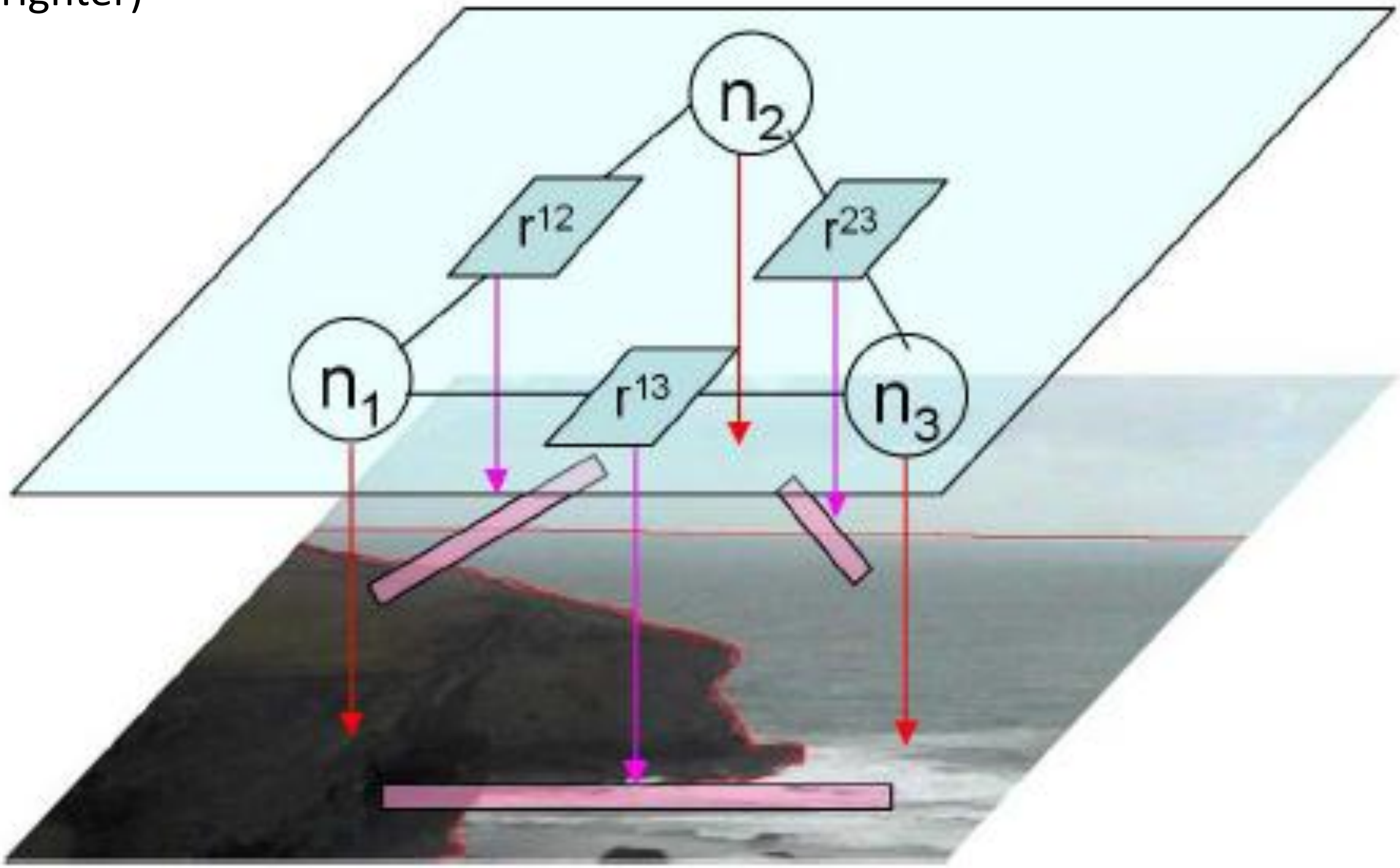
Image labeling (Loopy-BP)



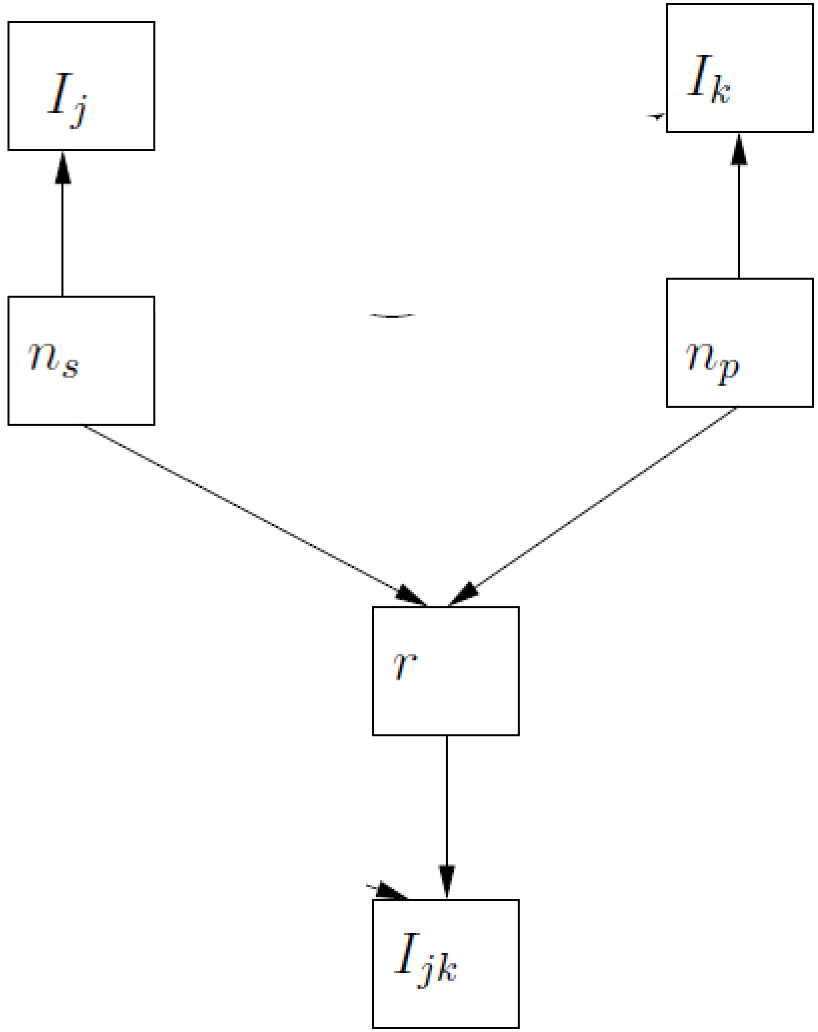
- *a* = class label (e.g., road, car, etc…)
- *b* = image data at local patch

Example from Carbonetto, de Freitas & Barnard, ECCV'04

Similar problems but using relations (Above, behind, below, left, right, beside, bluer, greener, nearer, smaller, larger, brighter)

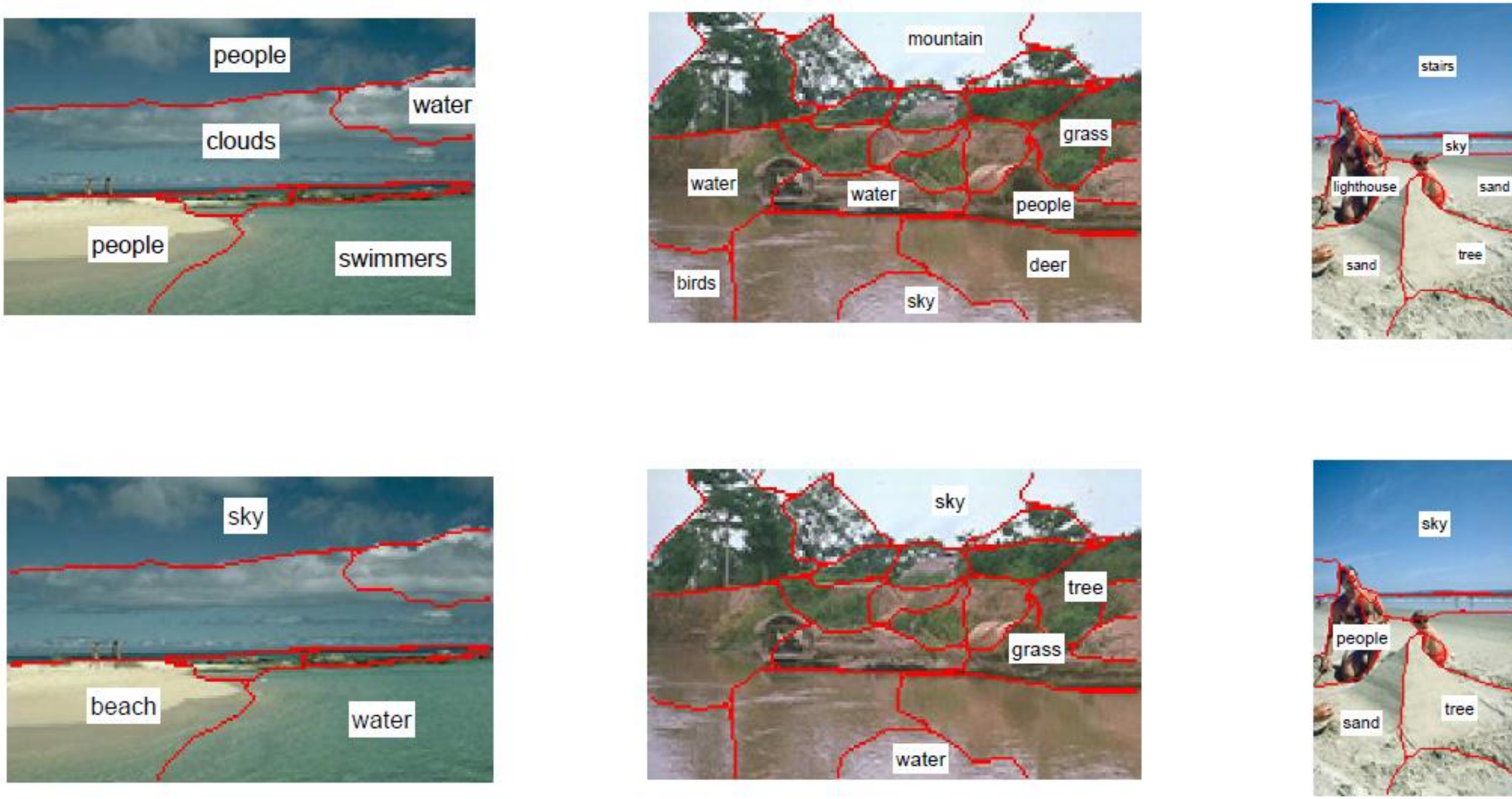


Abhinav Gupta and Larry S. Davis, Beyond Nouns: Exploiting prepositions and comparative adjectives for learning visual classifiers, ECCV 2008.

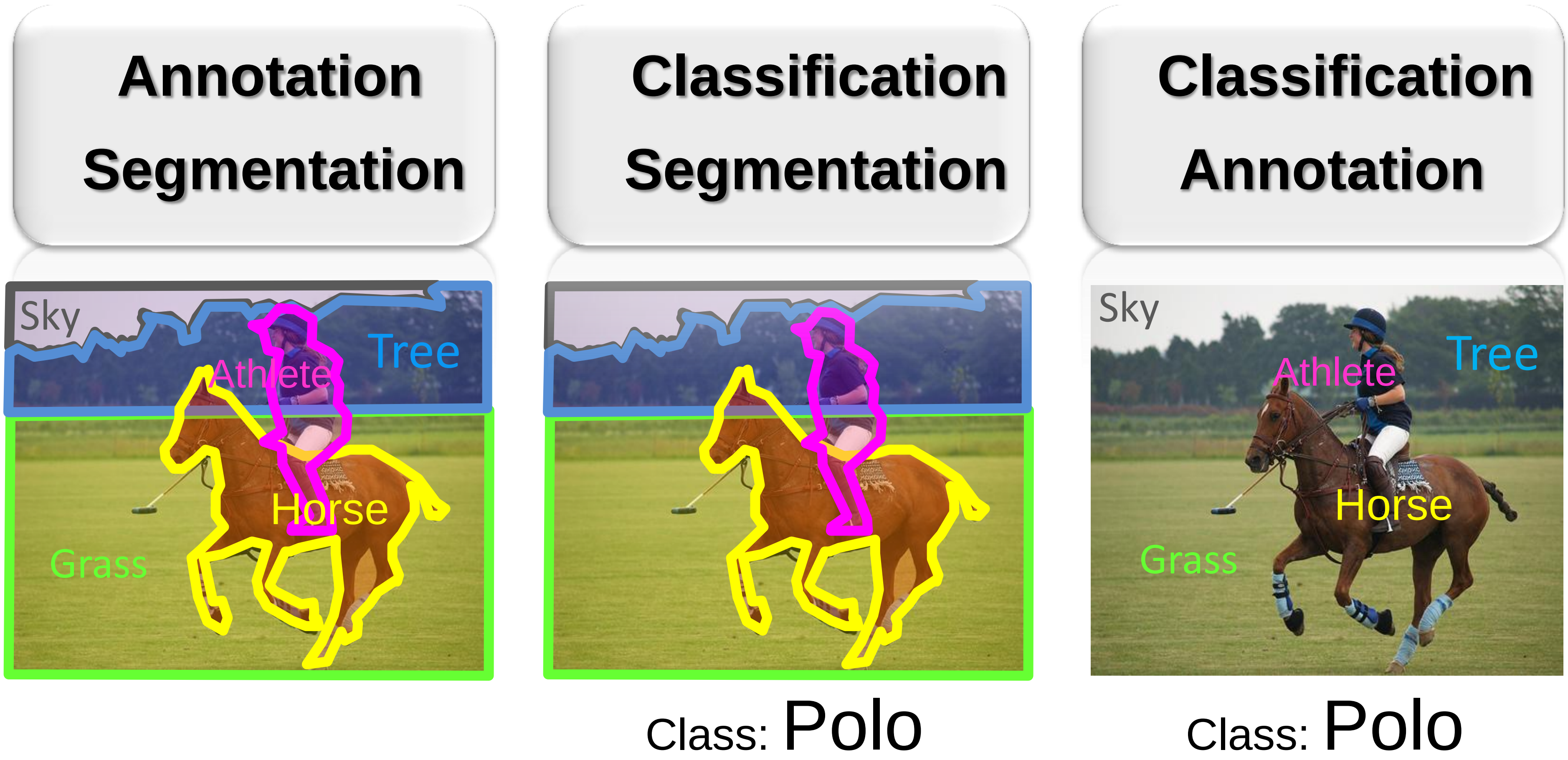


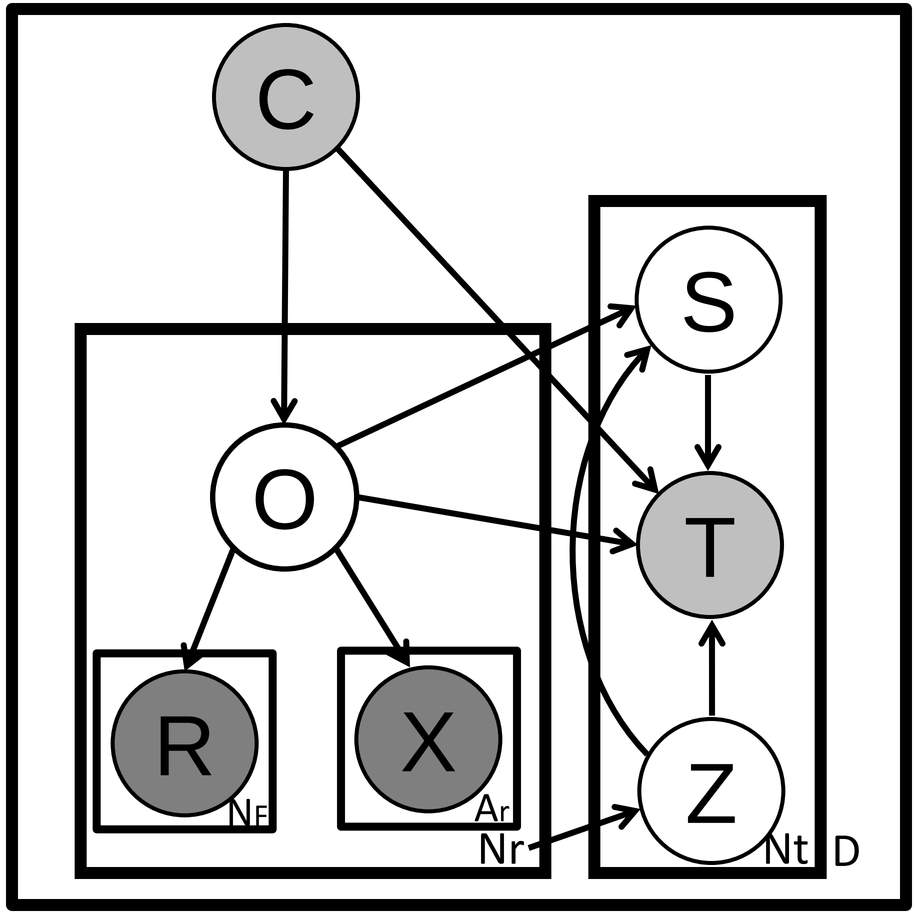
Inference: BP

$$P(n_1, n_2..|I_1, I_2..I_{12}, .., \quad) \propto \prod_i P(I_i|n_i \quad) \prod_{(j,k)} \sum_{r_{jk}} P(I_{jk}|r_{jk} \quad)P(r_{jk}|n_j, n_k)$$



people
water
clouds
people
swimmers
mountain
grass
water
water
people
birds
deer
sky
stairs
sky
lighthouse
sand
sand
tree
sky
beach
water
sky
tree
grass
water
sky
people
tree
sand

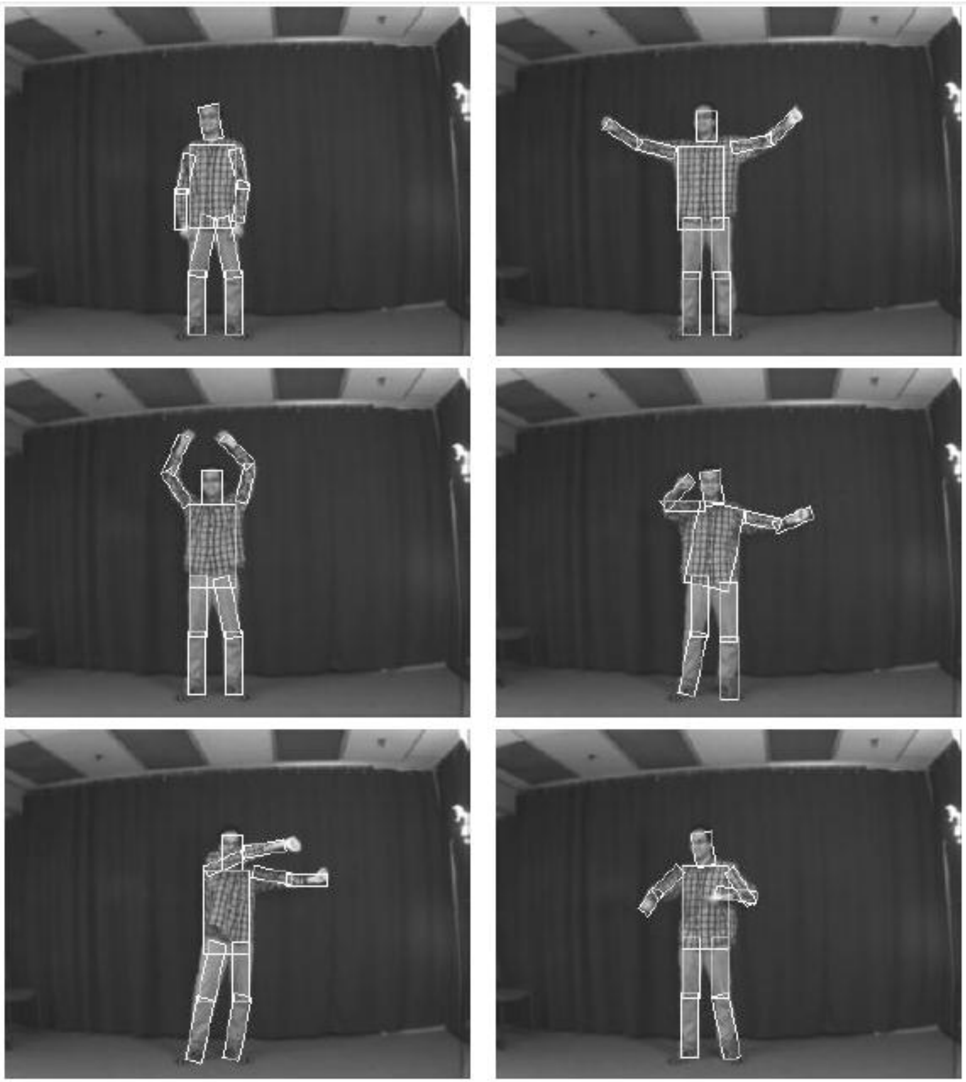

Annotation
Segmentation
Sky
Tree
Athlete
Horse
Grass
Classification
Segmentation
Class: Polo
Classification
Annotation
Sky
Tree
Athlete
Horse
Grass
Class: Polo



Athlete
Horse
Grass
Trees
Sky
Saddle

L.-J. Li, R. Socher and L. Fei-Fei. Towards Total Scene Understanding:Classification, Annotation and Segmentation in an Automatic Framework. CVPR2009

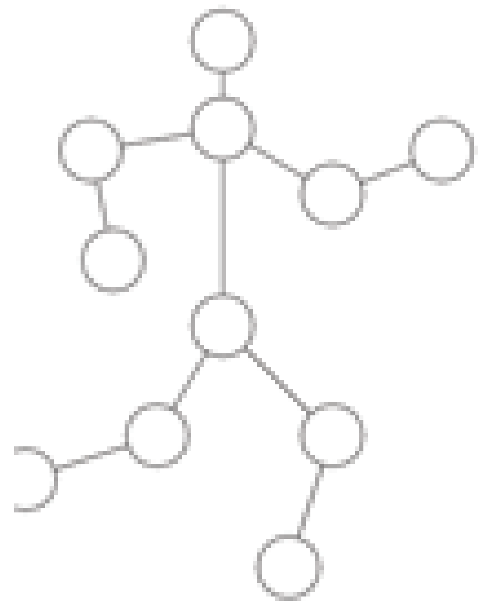
Example: Inferring human poses



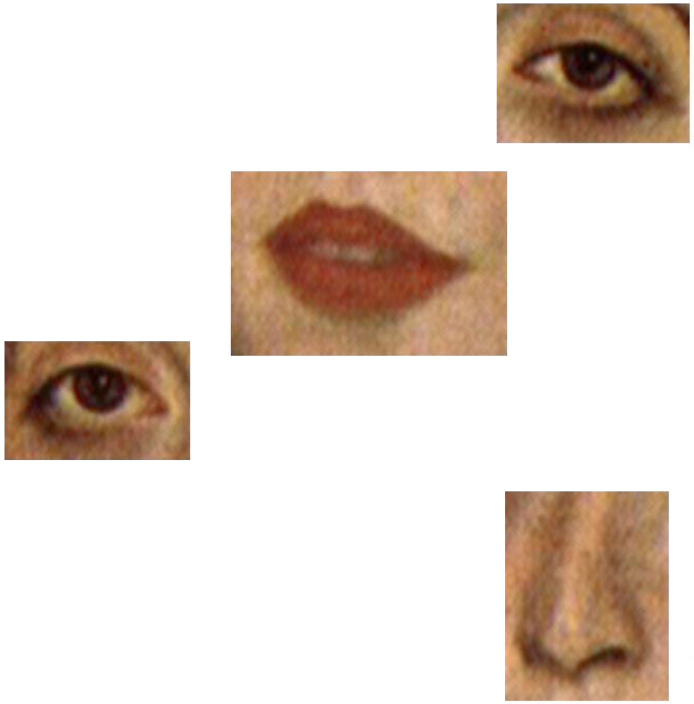
Example from Felzenszwalb'04

I= (features from) Input image dat

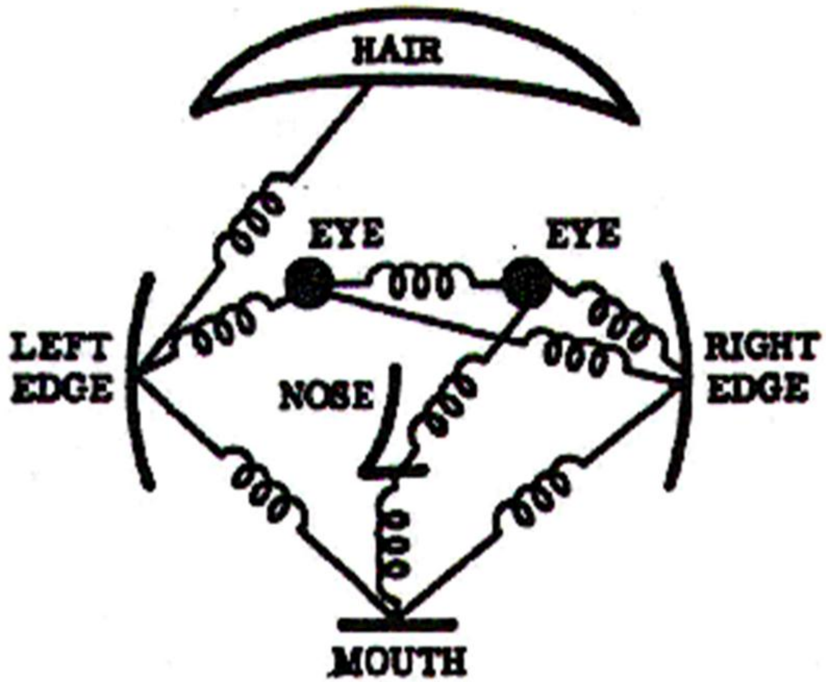
x_i = Pose (location and orientation) of limb i



General problem: Representing knowledge about spatial relations between variables



Is this a face?



[Fischler & Elschlager 73]

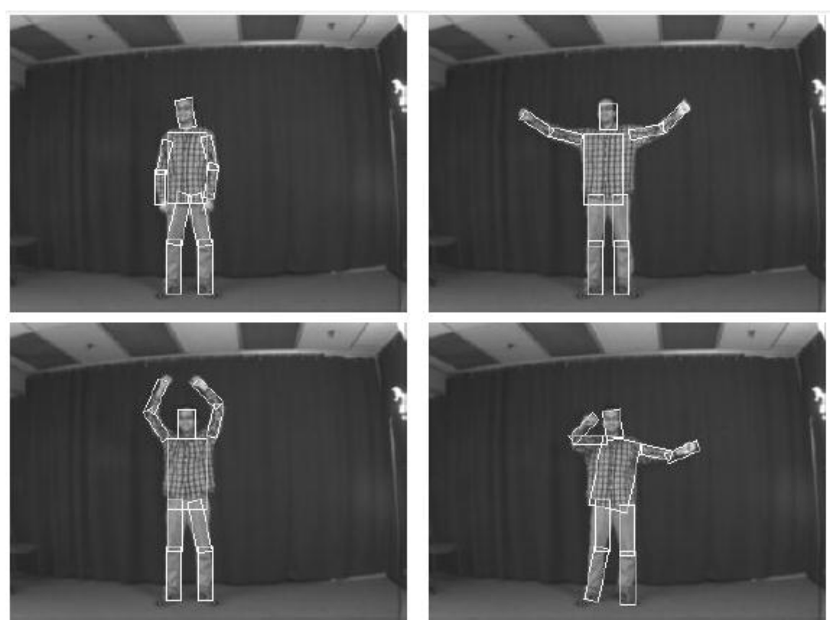
Example

$$P(X|I)\alpha \quad \prod_{i} \varphi_i(x_i, I) \prod_{i,j} \varphi_{ij}(x_i, x_j)$$

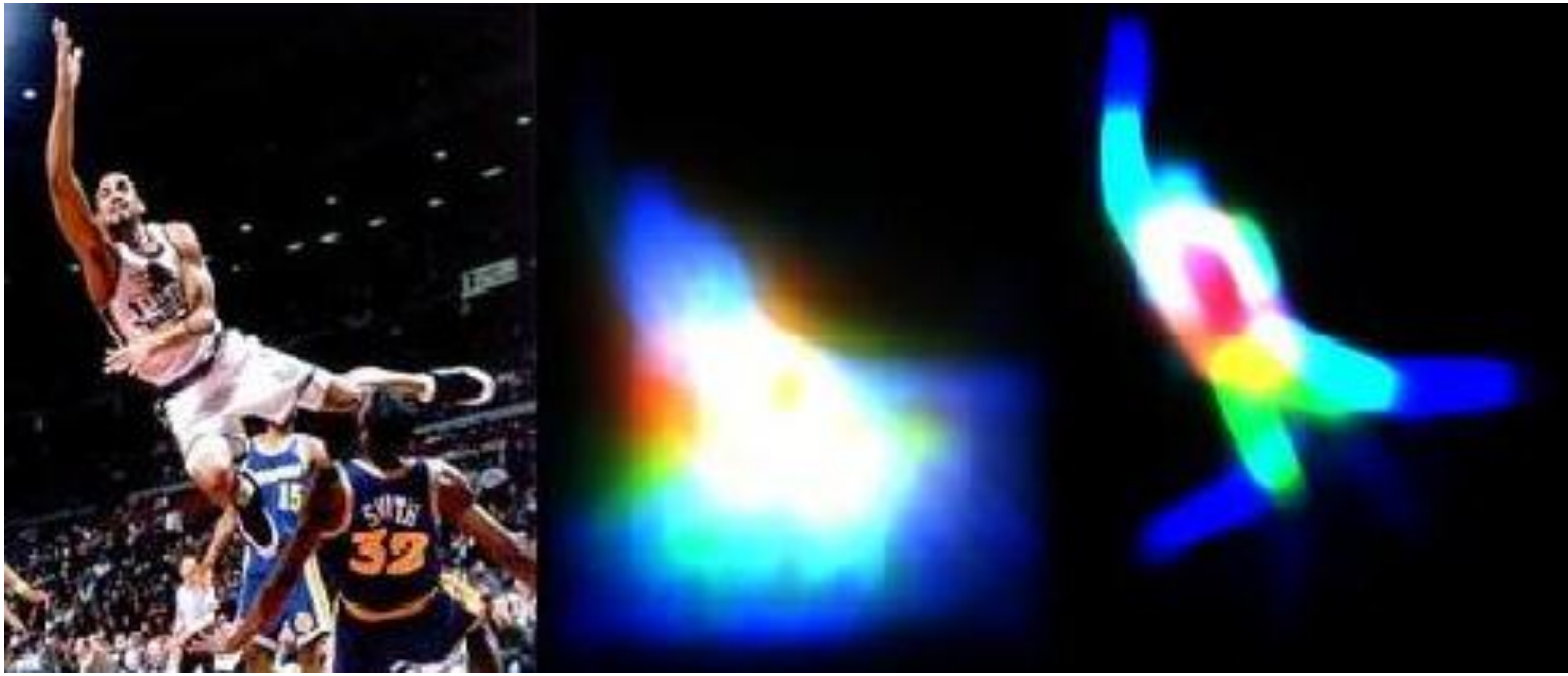
Agreement of location with image

Prior distribution of shapes

- Max-product: Tree-structure → DP algorithm, efficient
- Normally N^2 but reduction with Gaussian model for φ_{ij}



Sum-product marginals



•

D. Ramanan. Learning to parse images of articulated bodies. NIPS 2007.

Example

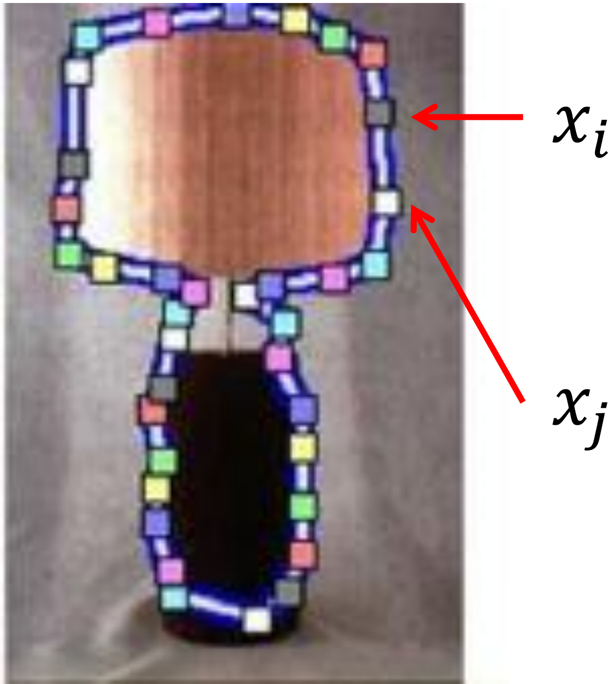
- Variables = locations of landmark points on shape (x_i) + measurements from image (I)
- Max-product inference

$$P(X|I) \alpha \prod_{i,j} \varphi_{ij}(x_i, x_j, \theta) \prod_{i} F_i(x_i, I) \prod_{i,j} F_{ij}(x_i, x_j, I)$$

Prior distribution of shapes

Agreement of location with image

Agreement line between 2 neighboring landmarks and image gradients



G. Heitz, G. Elidan, B. Packer, and D. Koller (2008). "Shape-Based Object Localization for Descriptive Classification."